Unsupervised Guitar String Classification for Tablature Transcription

Jonathan Michelson

April 2017

School of Music
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Richard Stern

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Music and Technology

For my dog

Abstract

Guitar tablature transcription (brief explanation) is a commonly used music notation standard that could benefit from automation. Previous work solves this task in a framework that requires prior information about the guitar. Here, an unsupervised offline solution is introduced. Inharmonicity estimates extracted from isolated guitar notes naturally segregate into semi-linear formations determined by its sourced string. The inharmonicities are modeled with a linear regression mixture, and optimized with expectation maximization.

Acknowledgments

My advisor is cool.

Contents

List of Figures

List of Tables

Chapter 1:   Introduction

The guitar is a widely popular musical instrument whose players run the gamut of expertise. Its most common types are classical, acoustic, and electric, which have different characteristic qualities such as body dimensions, fretboard dimensions, amplification medium, and string material. Its six strings' open tunings are usually $E2 \approx 82\text{Hz}$, $A2 \approx 110\text{Hz}$, $D3 \approx 147\text{Hz}$, $G3 \approx 196\text{Hz}$, $B3 \approx 247\text{Hz}$, and $E4 \approx 330\text{Hz}$, though the tuning and number of strings can vary. Sound is created by pressing strings onto the fretboard with one hand, and perturbing combinations of pressed and open strings with the other hand. Each fret increases the played string's pitch by one semitone, and fretboards are commonly inlaid with greater than twenty frets. Note that this extends each strings' pitch range well beyond its open-tuned neighbor.

Because the pitch ranges of a guitar's adjacent strings overlap, a given musical phrase can usually be realized in a few different ways. Conventional music scores represent passages as notes and chords, and therefore leave fretboard position ambiguous [see figure]. This can be problematic if one desires to uncover a virtuoso's intentional string-fret selections.

Tablature is an alternative music notation for guitarists that doesn't suffer from the one-to-many mapping of scores. Strings and fret positions throughout the piece are mapped pictorially to an ASCII representation of the guitar's six strings. (see figure). Because of its intuition lies in correspondence with the guitar's form rather than music theory, tablature is popular with self-taught students who have no prior musical training. And since simple text is the encoding mechanism, tablature is a lightweight portable medium. These features help explain why tablature is more popular than standard music notation (cite). A downside is its tedious creation process; tab creation usually means manual annotation of a guitar piece.

Automated tablature transcription provides a solution to tedious tab annotation, It can be used in music education for novice guitarists. It could also be used to uncover the fretboard positions of legendary guitar pieces, about whose tablature identities we can only speculate. In the literature, there are two main variants of transcription: a general version (in which goal is simply generation of a feasible tab), and transcription (in which goal is ascertainment of exact string-fret combos that produced audio).

What I intend to do (briefly): Introduce one unsupervised solution to the transcription problem. Specifically, we take inharmonicities and plot them, cluster them with expectation maximization for linear regression mixtures, allowing us to classify strings. Then with tuning information provided by the user, we can infer the tablature.

Outline of remaining sections: Discuss existing/related work in the generation and transcription field, detailed discussion of my idea, lay out experients, what i think they'll show, introduce evaluation crtieria: how well and when done.

Chapter 2:   Previous Work

This section introduces previous work on automated string classification or tablature transcription from strictly audio recordings. For visual, audio-visual, and instrument enhancement approaches to tablature transcription, the reader is referred to [10, 12]. [1]. We intentionally omit discussion of score-generation or transcription, though much work has been done in that area and we refer the reader to [—-] for more on that. [2, 4] [3]

Chapter 3: Proposed Solution

Our implemented system can be decomposed into the following high-level blocks: pre-processing, inharmonicity extraction, and classification. Each block is presented in detail below.

3.1 Pre-processing

In the pre-processing stage, onset detection and pitch estimation are used to extract guitar notes from the audio files and estimate their fundamentals, respectively. The proposed system operates only on monophonic isolated guitar recordings. Pitch detection in this setting is fairly trivial, and we found that harmonic sum spectrum [11] was straightforward and effective. In this technique, a signal's magnitude spectrum is repeatedly downsampled and summed, producing a composite spectrum that emphasizes harmonically-aligned peaks. Then the most prominent composite peak is usually selected as the f_0 candidate. In a variation known as harmonic product spectrum, the operation after downsampling is multiplication, which has the additional effect of de-emphasizing inharmonic spectral regions. equations.

Our solution Rectified spectral flux with pre-processing [5, 6]

3.2   Inharmonicity Estimation

3.2.1   Theory

When a guitar string is displaced by a pluck, its rigidness contributes to the restoring force so as to skew its harmonics upward in frequency [8] according to

$$f_k = k f_0 \sqrt{1 + \beta k^2}, \tag{3.1}$$

where f_0 is its fundamental frequency, f_k is its kth harmonic, and β is the amount of inharmonicity, given by

$$\beta = \frac{\pi^3 Q d^4}{64 T l^2} \tag{3.2}$$

As can be seen, the inharmonicity is a function of the material properties of the string: Q is Young's modulus, d is its diameter, T is its tension, and l is its vibrating length. Note that in equation 3.1 when $\beta = 0$, the kth harmonic's frequency reduces to the ideal relationship $f_k = k f_0$ since zero inharmonicity is present. Interestingly, this phenomenon is perceptible in the acoustic guitar in some cases (Karjalainene [cite]), but more importantly is key in the task of string discrimination.

To understand inharmonicity's discriminative potential, consider equation 3.2. If we make the simplifying assumption that terms Q, d, and T are constant for a given guitar string during a performance, then the inharmonicities β of each note produced by that string are simply a function of l^{-2}, the inverse of the square of each notes' vibrating lengths. But note that for the guitar, length and fret number, or equivalently pitch, are interchangeable, so ....

Inharmonicities for each string's notes are therefore restricted to a quadratic function of pitch specified by the string's unique material constants. When we plot inharmonicities versus MIDI note number, we see that the notes naturally segregate into semi-linear clusters indicative of the strings on which they were played. The semi-linearness comes from the logarithm used when

transforming pitches that are harmonically exponential to a harmonically linear scale, i.e. MIDI note numbers. This is the basis for our proposed solution in the next section.

To estimate the inharmonicity of a note, we follow a combination of Barbancho's [4] and Abesser's [1] methods. We first note that we can rearrange equation (3.1) as

$$(\frac{f_k}{f_0})^2 = k^2 + \beta k^4, \tag{3.3}$$

which expresses the partial-to-fundamental ratio in terms of the inharmonicity β. We can approximate this as a 4th-order polynomial

$$(\frac{f_k}{f_0})^2 \approx \sum_{i=0}^{4} p_i k^i, \tag{3.4}$$

which allows us to reasonably estimate β as

$$\beta \approx p_4, \tag{3.5}$$

the coefficient on the 4th order term in the partial-fundamental ratios' polynomial fit.

We follow Barbancho's iterative inharmonicity estimation method [4]. After a guitar note onset is detected,

1. Take three consecutive Hanning-windowed 100-ms frames and average their DFT magnitudes.

2. Estimate its f_0.

3. Initialize inharmonicity estimate $\beta = 0$ and list of number of partials to locate K, e.g. $K = \{10, 20, 40\}$.

4. Search the note's magnitude spectrum for the $k = \{1, 2, ...K\}$ inharmonic partials, whose locations we expect to be $f_k = k f_0 \sqrt{1 + \beta k^2}$.

5. Compute deviation vector $\mathbf{d}$, whose kth component d_k is the kth inharmonic partial's de-

viation from its corresponding ideal harmonic kf_0.

6. Fit polynomial $d = ak^3 + bk + c$ to deviations measured in previous step. Update estimate of the inharmonicity coefficient: $\beta \longleftarrow \frac{2a}{f_0+b}$.

7. Repeat steps 4 - 6 for each K value.

8. Return final estimate of the inharmonicity coefficient β.

3.2.2 Estimation

ACTUAL INHARMONICITY EXTRACTION ROUTINE: After a note onset is detected, we take three consecutive Hanning-windowed 100-ms frames and spectrally average their DFT magnitudes. Then we search the note's magnitude spectrum for K peaks close to integer multiples of the note's fundamental. We compute each of the K peaks' deviations from their corresponding ideal locations, and save them in a deviation vector $\mathbf{d}$, where each component $\mathbf{d_k}$ is the amount by which the kth located peak deviates from the ideal kth harmonic $f_k = kf_0$. Next, a 4th order polynomial is fit to the deviation curve according to the estimation discussion above.

3.3 Classification

After refined inharmonicity estimates of all notes are obtained, we plot them with respect to their fundamental frequencies. Doing so reveals that notes played on the same string cluster in semi-linear groups, consistent with the simplifying assumptions made in the theory discussion above. The data is amenable to a linear regression mixture model, in which one optimizes the parameters of k linear regressions to maximize the likelihoood..... [7]

3.3.1 Linear Regression Mixture Models

3.3.2 Expectation Maximization

$$y_i = \mathbf{w_j^T} \mathbf{x_i} + \epsilon_{i,j}, \pi_j \tag{3.6}$$

i is the observation index such that $\mathbf{x_i}$ is the ith guitar note and y_i is its inharmonicity, j is the jth latent linear regression mixture, $\mathbf{w_j}$ is the jth linear regression's weights, and $\epsilon_{i,j}$ is the Gaussian noise term such that $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma_j^2)$. π_j is the mixing probability of mixture j.

The parameters we iteratively update according to the EM framework are $\mathbf{w_j}, \sigma_j^2$, and π_j for each mixture j. These are found using: (EM equations here).

Chapter 4:   Preliminary Results

4.1   Dataset

The proposed solution was evaluated on the RWC dataset [9].

4.2   Experiments

4.3   Evaluation

Chapter 5:   Conclusion

Bibliography

[1] J. Abeßer. Automatic string detection for bass guitar and electric guitar. In *From Sounds to Music and Emotions - 9th International Symposium, CMMR 2012*, pages 333–352, London, UK, June 2012.

[2] A. Barbancho, A. Klapuri, L. Tardón, and Isabel Barbancho. Automatic transcription of guitar chords and fingering from audio. In *Proceedings of the IEEE Transactions on Speech and Language Processing*, pages 915–921, March 2012.

[3] I. Barbancho, A. Barbancho, L. Tardón, and S. Sammartino. Pitch and played string estimation in classic and acoustic guitars. In *Proceedings of the 126th Audio Engineering Society (AES) Convention*, Munich, Germany, 2009.

[4] I. Barbancho, L Tardón, S. Sammartino, and A. Barbancho. Inharmonicity-based method for the automatic generation of guitar tablature. In *Proceedings of the IEEE Transactions on Audio, Speech, and Language Processing*, pages 1857–1868, August 2012.

[5] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13 (5):1035–1047, 2005.

[6] S. Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.

[7] S. Faria and G. Soromenho. Fitting mixtures of linear regressions. *Journal of Statistical Computation and Simulation*, 80(2):201–225, February 2010.

[8] N. Fletcher and T. Rossing. *The Physics of Musical Instruments*. Springer-Verlag, 1998.

[9] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 229–230, October 2003.

[10] A. Hrybyk and Y. Kim. Combined audio and video for guitar chord identification. In *Proc. of the 11th International Society for Music Information Retrieval (ISMIR)*, pages 159–164, Utrecht, Netherlands, 2010.

[11] A. Noll. Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In *Computer Processing in Communications Proceedings*, pages 779–797, New York, N.Y., April 1969.

[12] P. O'Grady and S. Rickard. Automatic hexaphonic guitar transcription using non-negative constraints. In *Proc. IET Irish Signals Syst. Conf. (ISSC 2009)*, pages 1–6, Dublin, Ireland.